Image Angel Technical White Paper

Introduction

Image Angel protects images through on-demand digital forensic watermarking and detection. This whitepaper describes these technologies and how they are combined to provide the Image Angel service.

To enquire about applying Image Angel to your product, visit imageangel.co.uk
Download our server-side SDK from npmjs.com/package/image-angel
Source code and documentation are at github.com/image-angel/client

Digital Forensic Watermarking

A digital forensic watermark modifies an image to embed a marker and a small amount of information. This modification is entirely in the 'signal', i.e. the visible pixels of the image. It does not depend on image metadata - which can easily be read and modified. This modification to the image can later be detected and the information extracted and used to identify the origin of the image.

Watermarks may be blind or not. A non-blind watermark is only detectable if the original unmodified image is also available, by inspecting the differences between the original image and the suspected watermarked image. On the other hand, a non-blind watermark is detectable without having access to the original. Non-blind watermarks are considerably more challenging to design, since the watermarked image must be subtly different from any naturally-occuring image. However, they are much more convenient to use.

The design of a watermarking method involves making tradeoffs between various desirable properties:

Robustness - this is the extent to which the watermark is still readable after modifications have been made to the image such as cropping, screenshotting, or filtering.

Perceptibility - the extent to which the watermarked image looks different to the original image, and the perceived quality of the watermarked image.

Capacity - the amount of data that can be embedded in the watermark. Typically measured in bits of information.

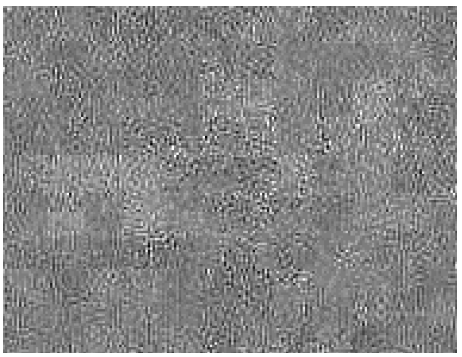
The ideal digital forensic watermark would be perfectly robust, entirely imperceptible, and have a very high capacity. This combination of properties is not in practice possible - increasing capacity means either reducing robustness or else making the watermark more perceptible. Likewise, imposing very strong perceptibility constraints means making the watermark either less robust because very subtle changes are being made which can be removed by filtering.

The Image Angel watermark makes the following design decisions:

- Semi-Blind - the original image is not needed to detect and read the watermark, but making it available may improve detection ability.
- Robust to screenshotting, resizing, and cropping up to one quarter of the image
- No perceptible loss of image quality.
- Minor perceptible difference when viewed side-by-side with the original image.
- 32-bit information capacity, sufficient to embed a unique transaction identifier.

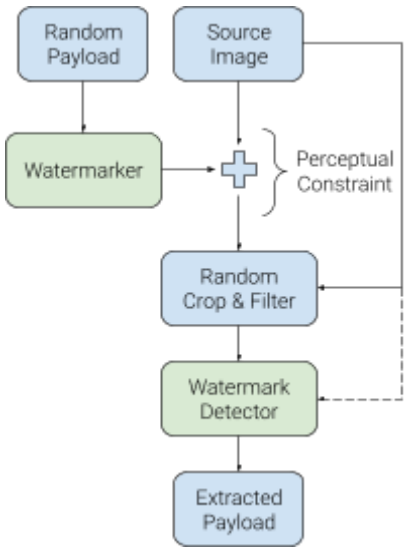
An example watermarked image. Original is on the left, watermarked is on the right.

The image below shows a zoomed-in and exaggerated version of the single-channel difference between original and watermarked images:



Difference in blue channel (using RGB colourspace)

We continue to research and improve the Image Angel watermark using AI to improve robustness and perceptual quality. The research pipeline looks like this:



The green *Watermarker* component is responsible for creating the watermark modulation which is applied to the image, subject to a perceptual constraint.

The green *Watermark Detector* is presented with cropped and filtered images which may or may not have had a watermark applied to them. It is responsible for determining whether or not the image had a watermark, and if so extracting the watermark payload.
It is tested both with and without access to the original unwatermarked source image.

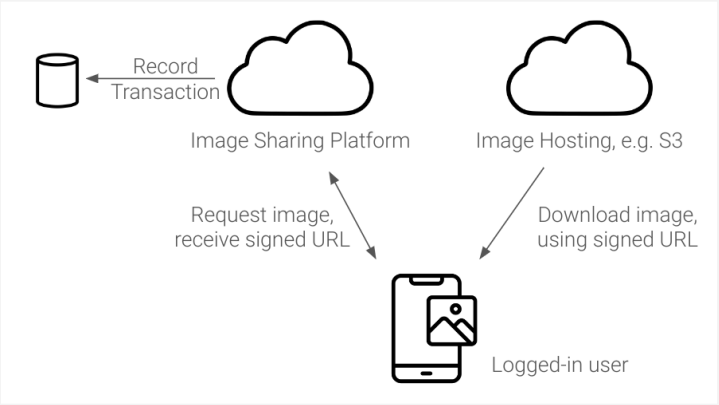
The overall objective function for these two components is the success rate at detecting and extracting watermarks, subject to a penalty for false detection, and a constraint on the perceptual difference that the watermark makes to the image.

Watermark Research Pipeline

We will roll out improvements to our watermarking algorithm from time to time. When we do this, we will continue to support detection of watermarks applied using prior versions of the watermarking algorithm for at least 12 months after the point of sharing. This is independent for each sharing event, so an old or archive image which is newly shared will continue to be protected for 12 months from that point.

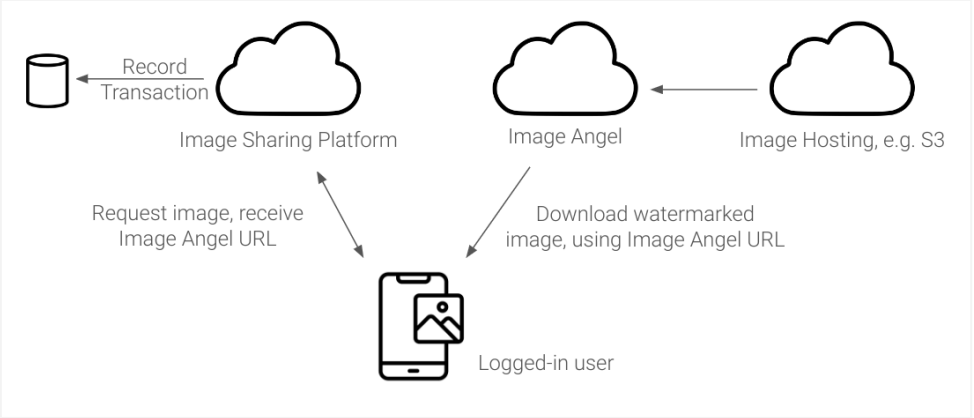
On-Demand Watermarking

Image Angel is designed to be as simple as possible to integrate into an image-sharing platform. When sending an image to a user without Image Angel, typically a URL will be supplied to the client device (phone or browser) which the device accesses in order to download the image. This URL might be for example a limited-lifetime signed URL for an AWS S3 bucket:



Example flow without Image Angel

With Image Angel, the Image Sharing Platform wraps the image download URL in an ImageAngel URL. This can be done locally on the image sharing platform server, with no need to contact Image Angel servers. This URL is supplied to the client, which then contacts Image Angel to obtain the individually-watermarked image:



Corresponding flow with Image Angel

Note that the use of S3 signed URLs here is just an example; any URL that is accessible by Image Angel can be used. This URL will not be revealed to the user, which means they are not able to access the unwatermarked image.

Image Angel may cache the watermarked version of the image for a short period of time, in case the user makes repeated downloads, for example if opening on a different device or refreshing a page.

Anatomy of an Image Angel URL

The Image Angel URLs which are constructed by the platform and supplied to the client have the following format:

```
https://api.imageangel.co.uk/wm/name.jpeg?u=auxinfo&s=AgV4N
8bbK90%2B3KHWgsZCiCTeKgizLJK6zs14kpMSL3lTKJgAkwAEAAFhAAJ2MA
AVYXdzLWNyeXB0by1wdWJsaWMta2V5AERBZ1RuYjBxSmNlSWtyVzdXcXNVS
Xpnb0VzYUUvUXR5eDJXMjFNWTgzSXExR0dJdndSa0I1UVJsYk9kR3gvSW5B
SlE9PQABdQAgYzQ5OWQzMmQ3ZjMzNjUyNzczYjc3MzBlZTZlZTAxNTAAAXc
AAzRkMgABAAtpbWFnZS1hbmdlbAAcdGVzdC1rZXkAAACAAAAADMSNMOYUih
WGY2NGPgAwr6ouO0Kcx4xashgrFtvVhFPrzbOS%2BD6AWDY7zGqfwieirw1
VV4RtWVsNmELdHz3WAgAAEAAl7DavKiv3EC4GcgqNo1vc2k%2BMz2LHS4FS
zS%2FCj2vsVgtthZUihpB5FHpGUTBz1Oz%2F%2F%2F%2F%2FAAAAAQAAAAA
AAAAAAAAAQAAACYO4KKUszW8WlX3o7OyebTGOw2zZPjBwLqqwgmVDf7CDn
F%2BfmzeM11sm6%2FiljDQEpbfxsRcRucAZjBkAjAQ9ie2Gr9zCKRQaFWUx
sUUs2PZHYsShw4NgoPfCvFi8pwhDaAlIVn4ZjRH7Ji4xYoCMDBevU2G0CHZ
MBG0qrNI8rVS5cL3tKPjJVvhvv%2FR1UK2hOAEitlo5qsE0olBWZ6vcQ%3D
%3D
```

Breaking this into pieces:

`https://api.imageangel.co.uk/wm/`	The Image Angel API endpoint for watermarking.
`name.jpeg`	The filename that the client will see for the downloaded image.
`?u=auxinfo`	Any additional information the platform wishes to embed in the URL. For example the user's name.
`&s=AgV4…`	The base64-encoded encrypted image download URL. This is encrypted using additional authenticated data (AAD) giving the watermark id to embed and the auxinfo present in the URL.

Encrypting the URL ensures that the user cannot simply download the unwatermarked image themselves. The use of encryption with AAD means that it's not possible for them to change the watermark id or the aux info without this being detected by Image Angel and the image request being rejected.

The encryption algorithm used is the AWS default for symmetric encryption: AES-256 in GCM mode with a 12 byte IV and a 16-bit authentication tag.

Watermark Detection

If a content creator finds one of their images in the wild, and wishes to take action as a result, they will get in touch with Image Angel, providing the leaked image. We will then extract the watermark, identify the platform that it originated on, and contact that platform on their behalf, providing the transaction id that was embedded in the watermark.

If the image has been extensively processed or cropped we might not be able to extract the watermark. In such cases, it may be helpful if the content creator can provide the original unwatermarked image for comparison purposes. We will try to avoid the necessity for this if possible.